Proteomics: delivering new routes to drug discovery – Part 2

Kewal K. Jain, Bläsiring 7, CH-4057 Basel, Switzerland, tel/fax: +41 61 692 4461, e-mail: jain@pharmabiotech.ch, web: http://pharmabiotech.ch

IBC's *Proteomics – Delivering New Routes to Drug Discovery* conference (14–17 May 2001, Philadelphia, PA, USA), provided broad coverage of the application of innovative proteomic technologies for drug discovery. The second part of this conference review deals with the analysis of proteomic information, and details of technologies are described elsewhere[1].

Analyzing the human proteome

The first part of this conference dealt with the role of proteomics in drug development and structural genomics[2]. Scott Patterson (Celera Genomics, Rockville, MD, USA) described proteomics using liquid chromatography–MS (LC–MS), as opposed to two-dimensional (2D) gel electrophoresis (GE), which offers higher throughput and simpler automation. Integration of genomic, proteomic, transcript-level and biological data is achieved using bioinformatics. Leigh Anderson (Large Scale Proteomics, Germantown, MD, USA) discussed the integration of protein chips with the existing technologies of 2D GE and MS. Protein chips are more sensitive than 2D GE or MS, have a wide dynamic range, high specificity and excellent quantification capability. The drawbacks are poor correlation with phenotype, and that mRNA does not reflect protein amount.

Whereas classical 2D electrophoresis (2DE) has definite limitations with respect to resolution, reproducibility and protein-loading capacity, 2DE using immobilized pH gradients (IPGs) has proved to be amazingly flexible in terms of the requirements of proteome analysis[3]. Angelika Görg (Technical University of Munich, Freising-Weihenstephan, Germany) described the advances in analytical and micropreparative 2DE of proteins using wide-range IPGs up to pH 12.0 for overview patterns, as well as narrow, overlapping IPGs (between pH 2.5 and pH 12.0) with extended separation distances for the detection of a maximum of spot numbers, avoiding the presence of multiple proteins per spot and/or cross contamination of neighbouring spots for further investigation.

Chris Sutton (Kratos Analytical, Manchester, UK) described matrix-assisted laser desorption ionization (MALDI) as an essential tool for current proteomics, with applications in high-throughput protein identification and biomarker identification. The next generation of MALDI instruments is being designed for the structural elucidation of protein sequences and post-translational modifications to enable our understanding of biological function. Andreas Köpke (WITA Proteomics, Teltow/Berlin, Germany) described a high resolution 'all-in-one' system (both dimensions in one chamber without gel handling) of 2D GE for diagnostics and drug targeting. Non-equilibrium-pH gel electrophoresis (NEPHGE) can be applied within a pH range of 2–11 to samples containing 25–100 μg protein or <1 mg of tissue, and can separate up to 10,000 spots. An MS platform consisting of MALDI-MS and electrospray ionization (ESI)-MS–MS-Q-TOF (quadrupole time-of-flight) has been used for the detection of lung metastases in clinical trials, in an effort to determine the effectiveness of a compound in suppressing lung cancer. An assay could detect metastases consisting of <2000 cancer cells.

Protein biochips and microarrays

In understanding protein structure–function relationships, a knowledge of binding stability, multi-molecular complex assembly and crucial amino acid residues is essential. JoAnne Bruno (Biacore, Piscataway, NJ, USA) presented a chip-based, surface plasmon resonance (SPR) approach to monitoring protein–protein interactions. This method has the advantage that measurements are label-free and in real-time and, in many cases, samples do not require pretreatment before they are analyzed[4]. This powerful approach not only identifies binding partners to target molecules, but also provides binding specificity, active concentration, and quantitative affinity and kinetic data. SPR is useful for screening for ligands to orphan receptors, identifying functionally active proteins, characterizing protein–protein interactions, and exploring protein-complex assembly.

It is essential to develop analytical tools capable of systematically reporting on the output of each and every gene in the human genome. In the absence of hybridization assays or a PCR-equivalent, high affinity and high specificity antibodies represent the most logical path to success. Antibody production is rendered genomically relevant through parallel generation and screening, so as to reduce the time and cost associated with high quality antibodies destined for incorporation on arrays. Ian Humphery-Smith (Universiteit Utrecht, Utrecht, The Netherlands) described peptide, protein and antibody arrays as an integrated approach to following human tissues in health and disease. However, many genes are difficult to clone or are associated

1359-6446/01/$ – see front matter PII: S1359-6446(01)01912-2

with problematic proteins. Therefore, an essential element of this procedure is a closure strategy centred on 'signature peptides'. Other aspects already integrated in this process include crucially important surface chemistries, expression-vector design, robotics and three-dimensional (3D) biocomputing for data management and pattern recognition.

Bioinformatics for proteomics

The use of protein and expressed sequence tag (EST) similarity to help gene prediction is widespread. ESTs are an invaluable resource for protein identification and characterization in proteomics. They also allow proteins to be identified in the absence of genome sequence data. Jeanette Schmidt (Incyte Genomics, Palo Alto, CA, USA) discussed the role of bioinformatics in organizing EST databases into full-length genes. More recent tools include GeneWise™ (Compugen, Richmond Hill, Ontario, Canada), which considers every possible gene prediction in a genomic sequence and compares each one with the protein profile. The best combined score of both the gene prediction and the protein profile is used to provide a simultaneous gene prediction and protein alignment[5].

New tools have become available to analyze, search and store 2D-gel images in databases using differential protein-expression patterns. Suzanne Mattingly (Scimagix, Redwood Shores, CA, USA) described how new capabilities of 2D-gel-analysis software and 2D-gel databases can be leveraged for drug discovery research. Recent advances in the analysis and retrieval of 2D gels have included the automated registration of gels into an Oracle™ database, thus enabling rapid search and retrieval. A data-reduction capability can improve the recognition of patterns. Advantages of database storage of 2D-gel data include: options to quickly process data from prospective studies; rapid evaluation of results to discover new proteins; decisions for early termination of a compound with undesirable characteristics; and switching to more promising options. The recognition of protein-expression patterns facilitates the understanding of the mechanism of action of drugs. This approach can accelerate the drug discovery process.

Marc Wilkins (Proteome Systems, Sydney, Australia) described a bioinformatic platform for the integration and analysis of proteomic data to derive meaningful results. Challenges presented by the application of bioinformatics to high-throughput proteomics include the massive amount of data (>500 gigabytes per year) being generated, and the complexity of the data such as that found in protein-expression maps. Projects need laboratory information management systems (LIMS) and constant updating of bioinformatics tools in-house. Some measures to address these bottlenecks are complete process automation and client-server delivery via a web browser and integrated image analysis. When EST sequences are used for protein identification, they are usually first processed into contigs to reduce redundancy and generate longer sequences from the overlapping ESTs. However, the process of generating contigs can accidentally group biologically meaningful isoforms together. Protein identification and characterization of the dbEST database can be done by integrating clustering, translations, BLASTN and peptide-mass fingerprinting[6].

A variety of sequence and structural analysis methods are available to help assign biochemical function to proteins. Kim Fechtel (Genome Therapeutics, Waltham, MA, USA) described a method for the automated analysis of genomic and protein sequence. The PathoGenome™ data analysis process carries out BLAST analysis of all public genomic data and annotates data from other databases. The master catalog contains curated data compiled into >50,000 evolutionary families, as well as composite gene-product modules. Each family is associated with precalculated phylogenetic trees and secondary structure motifs. These comprehensive gene or protein databases enable rapid cross-referenced predicted functional analysis, without the need for bioinformatic computation by the user.

Applications of proteomics

Applications of proteomics in basic research, drug discovery, profiling drug effects, molecular diagnostics and various therapeutic areas were discussed. Stephen Pennington (University of Liverpool, Liverpool, UK) described the integration of proteomics into basic biomedical science. With mRNA analysis directing proteomic strategy, protein-expression mapping is considered to be a viable alternative to 2D electrophoresis[7]. This approach has been used to study the role of calcium signalling in the regulation of gene expression during cell-cycle progression.

Heart disease

In the post-genome era, proteomics will play an important role in the characterization of the global alterations in protein expression that underlies heart disease. The proteomic approach complements genomics-based and more traditional systems-based approaches, and will provide new insights into the cellular mechanisms involved in cardiac dysfunction. This improved understanding of the molecular basis of cardiac dysfunction should result in the development of novel therapeutics for the treatment of heart disease and chronic heart failure. Michael Dunn (Imperial College of Science, Technology and Medicine, Harefield, UK) presented the results of his studies on dilated cardiomyopathy (DCM), a disease of unknown etiology. The results of these studies have shown that the expression of some 100 cardiac proteins is altered significantly in DCM, with the majority of these proteins found to be missing in the diseased heart.

Problems of investigation of human material include disease heterogeneity, changes specific to disease and superimposed changes caused by heart failure and drug treatment. One approach to the investigation of this problem is to apply proteomics to appropriate animal models of human disease. Ubiquitin C-terminal hydrolase is greatly increased in both human and bovine DCM at the level of mRNA, as well as the expressed protein. The ubiquitin-proteasome pathway might be a possible target for therapeutic intervention in heart disease.

Diabetes

Stephen Fey (University of Southern Denmark, Odense, Denmark) demonstrated the use of proteome analysis in studying insulin-dependent diabetes mellitus. In rat models of the disease, the pancreatic islets were challenged by cytokines, which strongly regulated protein expression (up or down). These islets were grafted into diabetes-prone mice, re-isolated, labelled and analyzed by 2DE, which confirmed the earlier results. The selected proteins were identified by MALDI-MS. These studies are now being extended to human pancreatic islet cells. Although the genes involved in humans are different, the protein pathways involved are the same and play an important role in the genesis of the diabetes.

Infection biology

Peter Jungblut (Max Planck Institute of Infection Biology, Berlin, Germany) discussed the role of proteomics as a tool for infection biology. Several immunologically relevant proteins have been detected by comparing virulent strains with attenuated strains, studying secreted proteins, outer-surface proteins and analyzing the immunoproteome. Several vaccine candidates for *Mycobacterium tuberculosis*[8] and antigens for *Heliobacter pylori*[9] have been identified by proteomic approaches.

Toxicology and drug-safety assessment

Proteomics has been applied to toxicology and drug-safety assessment. Various studies have been conducted to study toxicity mechanisms and to assess the safety of new drug candidates by comparison with fingerprints of reference compounds with known toxicity. The molecular fingerprint of a drug is its gene-regulation pattern in response to the perturbances evoked by drug action and visualized by gene-expression profiling at the mRNA or protein level.

Sandra Steiner (Large Scale Proteomics) presented several projects to illustrate the use of proteomics to enhance the predictive power of safety assessment and accelerate the drug development process. Various preclinical markers have been used to monitor therapeutic versus toxicity mechanisms. An example is the effect of HMG CoA reductase inhibitors on cholesterol metabolism. This takes place through a tightly regulated pathway affecting a small number of proteins, and is upregulable by cholestyramine and lovastatin. The pattern of gene-network regulation induced in hepatic proteins as a response to lovastatin treatment was analyzed by proteomics in the livers of rats[10]. Lovastatin administration to the rats was associated with signs of toxicity, as reflected by changes in a heterogeneous set of cellular-stress proteins involved in functions such as cytoskeletal structure, calcium homeostasis, protease inhibition, cell signalling or apoptosis[10]. These results present new insights into liver gene-network regulations induced by lovastatin, and illustrate a yet to be explored application of proteomics to discover new targets by analysis of existing drugs and the pathways that they regulate.

Research and development

Paul Cutler (GlaxoSmithKline, Harlow, UK) discussed the role of proteomics in pharmaceutical research and development. The applications include clinical or tumour markers, biodistribution studies, toxicology, target identification, post-translational modifications and study of response to experimental treatment. An example was given of the MALDI-TOF MS of serum samples of transgenic mice expressing the human mutant ApoE*3-Leiden protein. ApoE*3 is a variant form of ApoE that shows reduced affinity for the low-density lipoprotein (LDL) receptor, and results in the dominant expression of type III hyperlipoproteinemia. This proteomic approach has enabled the ApoE*3-Leiden variant to be positioned in a 2DE separation of serum proteins, and has identified changes in the expression of haptoglobin, indicating that this protein could be a marker for the potential onset of atherosclerosis[11].

Concluding remarks

This was the most comprehensive proteomics conference that the author had ever attended. The calibre of the presentations was high, with the participation of academic proteomics centres. In addition to the presentations discussed, there were parallel sessions on sample preparation and strategic corporate alliances, as well as ample opportunities for audience–faculty interactions. The size of the conference reflects the expansion in proteomic technologies and applications.

References

1 Jain, K.K. (2001) *Proteomics: technologies and commercial opportunities* (7th edn) Jain PharmaBiotech Publications, Basel, Switzerland
2 Jain, K.K. (2001) Proteomics: delivering new routes to drug discovery – Part 1. *Drug Discov. Today* 6, 772–774
3 Görg, A. *et al.* (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21, 1037–1053
4 Nelson, R.W. *et al.* (2000) Biosensor chip mass spectrometry: a chip-based proteomics approach. *Electrophoresis* 21, 1155–1163
5 Birney, E. and Durbin, R. (2000) Using GeneWise™ in the *Drosophila* annotation experiment. *Genome Res.* 10, 547–548
6 Lisacek, F.C. *et al.* (2001) Strategy for protein isoform identification from expressed sequence tags and its application to peptide mass fingerprinting. *Electrophoresis* 22, 186–193

7 Jenkins, R. and Pennington, S.R. (2001) Protein expression mapping: Towards a viable alternative to 2-DE. *Proteomics* 1, 13–29
8 Mattow, J. *et al.* (2001) Identification of acidic, low molecular mass proteins of *Mycobacterium tuberculosis* strain H37Rv by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics* 1, 494–507
9 Bumann, D. *et al.* (2001) Proteome analysis of the common human pathogen *Helicobacter pylori*. *Proteomics* 1, 473–479
10 Steiner, S. *et al.* (2000) Proteomics to display lovastatin-induced protein and pathway regulation in rat liver. *Electrophoresis* 21, 2129–2137
11 Skehel, J.M. *et al.* (2000) Phenotyping apolipoprotein E*3-Leiden transgenic mice by two-dimensional polyacrylamide gel electrophoresis and mass spectrometric identification. *Electrophoresis* 21, 2540–2545

High content screening – from cells to data to knowledge

Catherine Liptrot, Senior Scientist, AstraZeneca, Cancer Research Area, 3F30 Mereside, Alderley Park, Macclesfield, Cheshire, UK SK10 4TG, tel: +44 1625 230147, fax: +44 1625 516033, e-mail: Catherine.Liptrot@astrazeneca.com

The second annual *IBC High Content Screening Symposium* (11–13 June 2001, Miami Beach, FL, USA) was sponsored by Cellomics (Pittsburgh, PA, USA) and Beckman Coulter (Fullerton, CA, USA) and brought together scientists from academia and industry with interests in cell-based screening. Presentations covered the capabilities of high content screening (HCS), its existing and potential applications, and how it can be improved in the future. In particular, there was significant input from Cellomics application scientists regarding ArrayScan II™ (AS; Cellomics) who discussed the different applications of this technology platform, in addition to product demonstrations and interactive discussions.

HCS versus HTS

Several companies are now providing technology that enables scientists to make multiple measurements of the cellular phenotype on a cell-by-cell basis. Compared with single-measurement well-based assays, multiparameter cell-based HCS assays yield data with much higher biological information content. Whereas HTS is used as a fast primary screen to identify hits for further testing, HCS can be used to identify leads from hits. The ability to obtain quantitative data from multiple endpoints, from both individual cells and cell populations, greatly enhances the information obtained from whole-cell screens in drug discovery. This could provide new insights into cell function and mode-of-action studies that were previously labour intensive and difficult to interpret.

ArrayScan™ technology

Keith Olson (Cellomics) reviewed the new software of the AS system, which enhances user interactions and improves assay throughput. The addition of Image-Pro™ software places greater flexibility for image-analysis algorithm development in the hands of the end user. The sub-population analysis feature enables users to select or reject cells based on nuclear size, shape, and intensity, or to gate on (that is, select a subpopulation of) cells that are above, below or between thresholds for a particular marker. Responders or non-responders to a stimulus (e.g. a cytokine) can be analyzed separately and compared with the entire cell population.

The advantages of the AS technology highlighted were: (1) images are acquired by a cooled CCD camera via standard inverted microscope optics (for most HCS applications the resolution is sufficient so three-dimensional detail obtained by laser-confocal approaches is unnecessary); (2) spatial information on individual cells (e.g. changes in cell shape or size), organelles (e.g. nuclear morphology) or cellular targets (e.g. transcription factor translocation to nucleus) can be obtained; (3) experiments are performed in standard clear-bottom 96- or 384-well microtitre plates; (4) cells can be live or fixed; (5) AS uses a UV arc lamp source, therefore, it can simultaneously quantify molecules that fluoresce over a wide spectral range.

Bright future for fluorescence technology

Fluorescent markers that are currently used by cell biologists could be adapted for use on the AS platform, according to Richard Haugland (Molecular Probes, Eugene, OR, USA). These include markers for organelles, nitric oxide and reactive oxygen species, intracellular Ca^{2+} (for G-protein-coupled receptor activation), enzyme substrates, fluorescent resonance energy-transfer (FRET) and enzyme-coupled assays. Haugland also described a

1359-6446/01/$ – see front matter PII: S1359-6446(01)01927-4